



# CHAT-PUB

Helper of welfare policy for youth

김강산  
양승빈  
박진호  
전창민

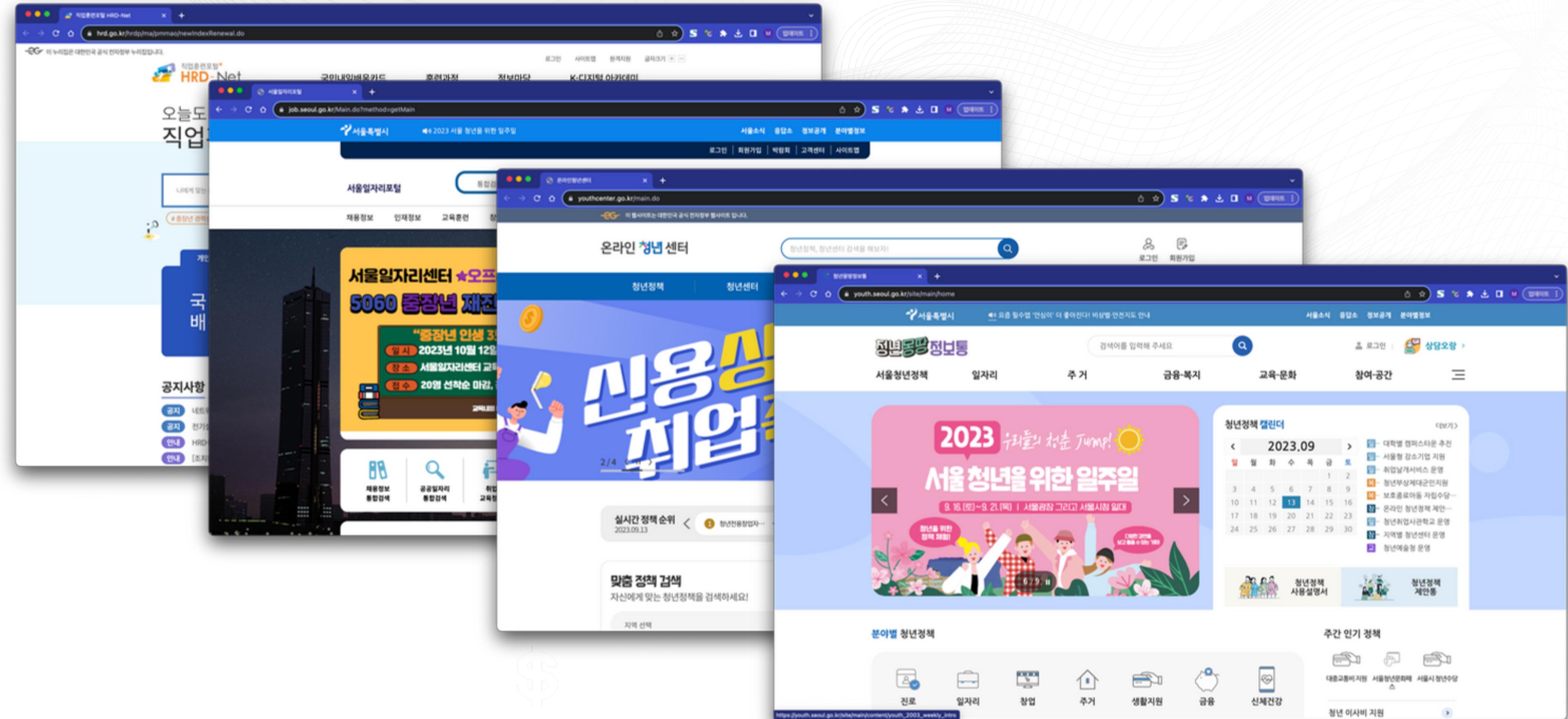


# TABLE OF CONTENTS

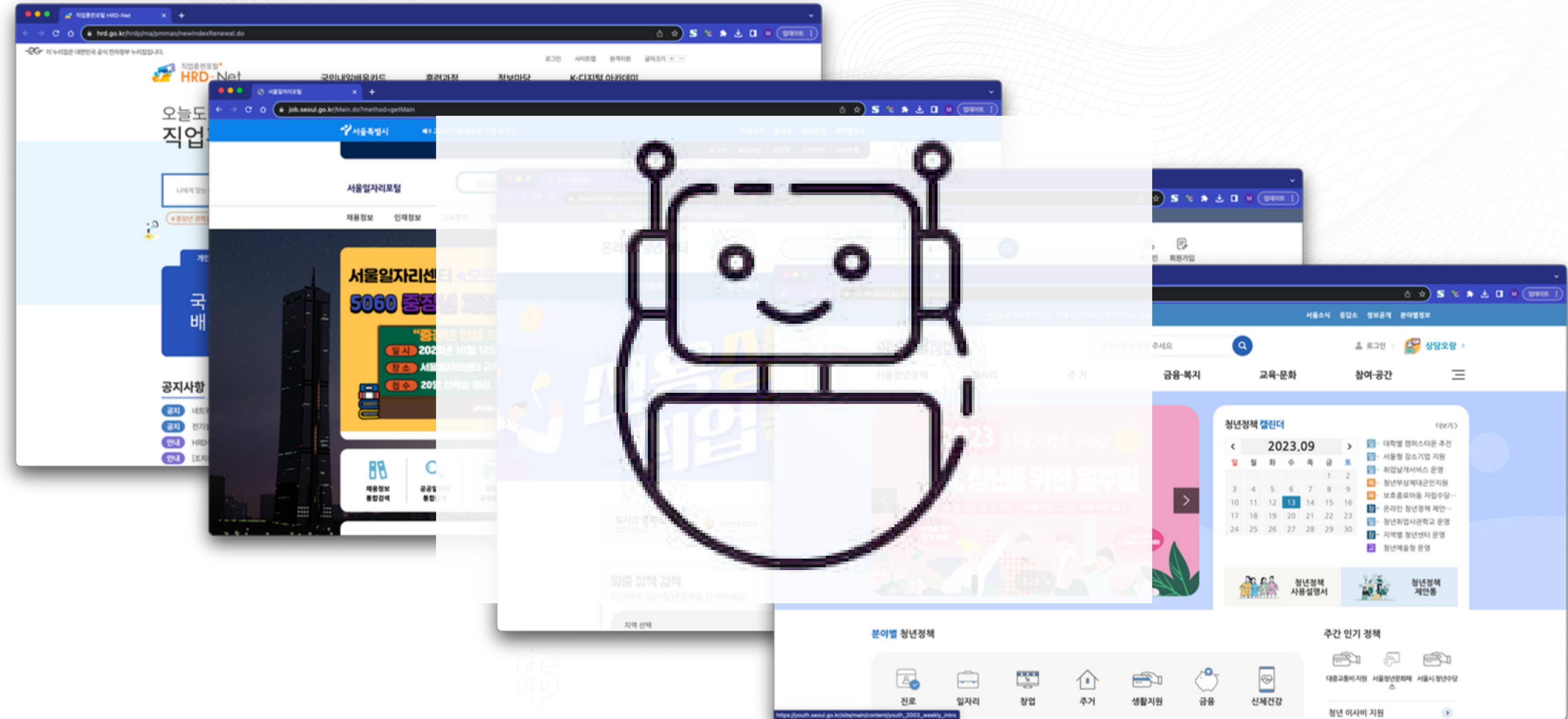
-  **Introduction**
-  **Role of each member**
-  **Implementation & Challenges**
  - Model
  - Back-End
  - Front-End
-  **Limitation**
-  **Schedule**



# INTRODUCTION



# INTRODUCTION



# INTRODUCTION

Our visions...



## Primary Goal

- A policy recommendation service that allows information on youth policies
- Create a highly accurate generative Chatbot using the NLP model
- Targeting young people living in Seoul and Gyeonggi Province



## Secondary Goal

- A rating system community
- Develop dedicated website
- Expanding service coverage to the metropolitan areas



## Last Goal

- A more detailed guide to the policy you want to proceed with
- Introduction to additional information required during the application process
- Nationwide service

# ROLE OF EACH MEMBER

## ➤ Kangsan Kim, Team Leader, UI/UX

- Designing better User Experience
- Visual Design
- Developing Web Page
- Figma, Nextjs

## ➤ Jinho Park, Data Engineer

- Data Pipeline
  - Crawling/Preprocessing/Generation ...
- Database Integration
- Code Refactoring
- Python, DBeaver

## ➤ Changmin Jun, Back-End Developer

- Login and Sign up with security
- Desgin and link DB
- Server Development
- Nextjs, MariaDB, Fast API

## ➤ Seungbin Yang, AI Engineer

- Model Development
- Model Training & Evaluation
- Paper Search
- Python, Faiss



- Designing better User Experience
- Visual Design
- Developing Web Page
- Figma, Nextjs

Jinho Park, Data Engineer

- Data Pipeline

- Crawling/Preprocessing/generation ...

Database Integration

Code Refactor

Python, DBeaver



Changmin Jun, Back-End Developer



GitHub

- Login and Sign up with security
- Desgin and link API
- Server Development
- Nextjs, MariaDB, PostgreSQL

Seungbin Yang, AI Engineer

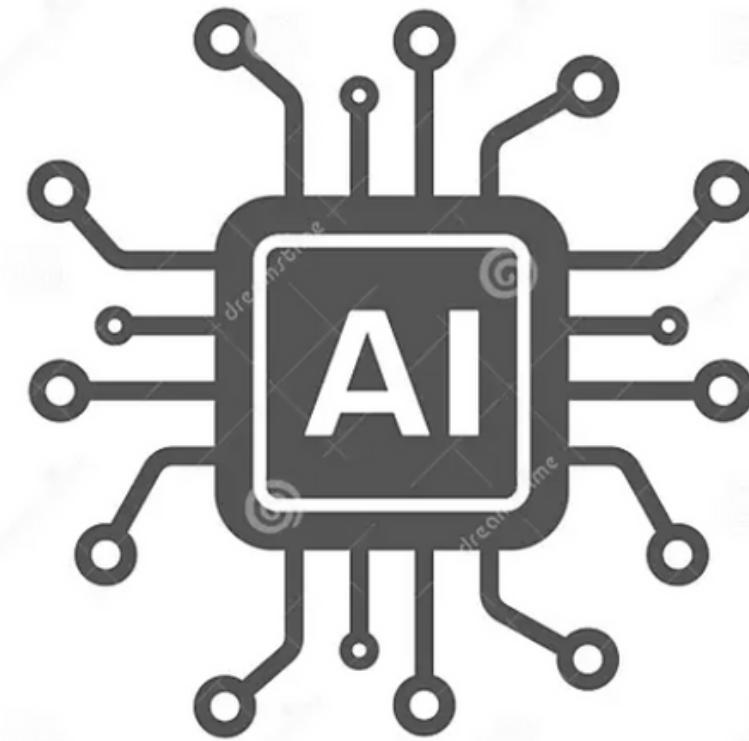
- Model Development

• Model Training & Evaluation

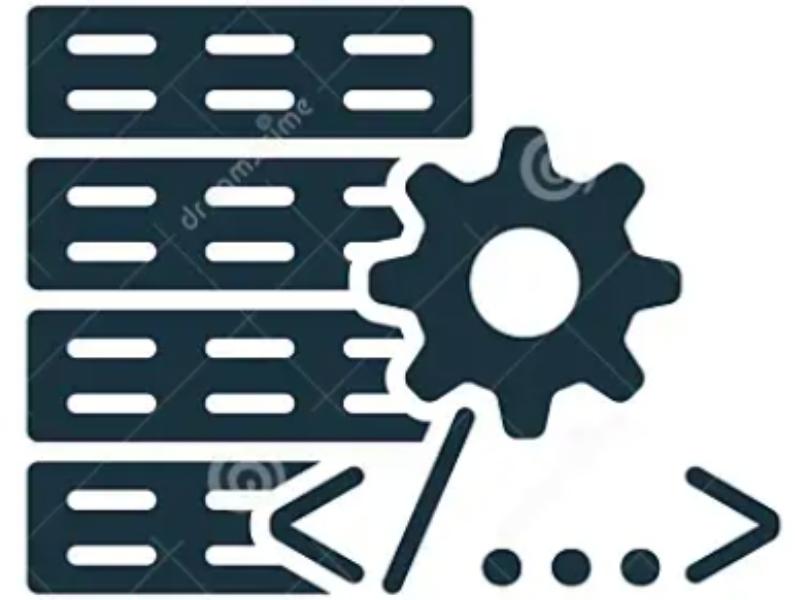
• Paper Search

• Python, Faiss

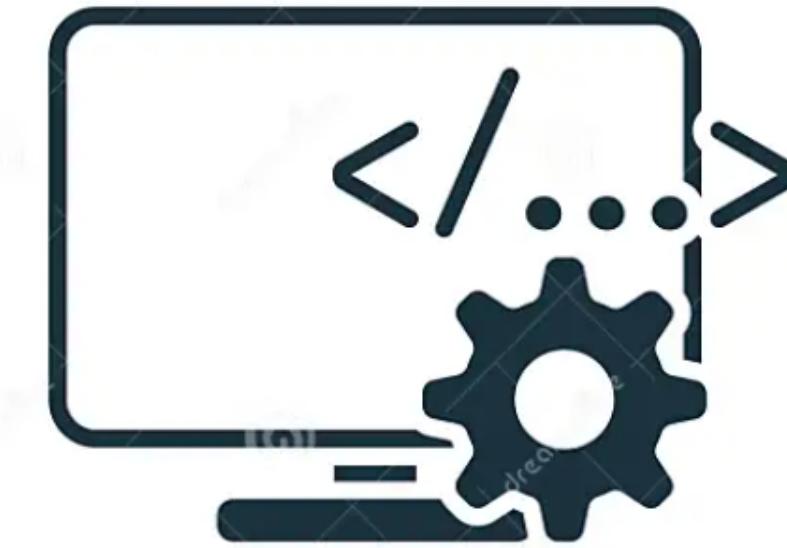
# IMPLEMENTATION & CHALLENGES



 **Model**



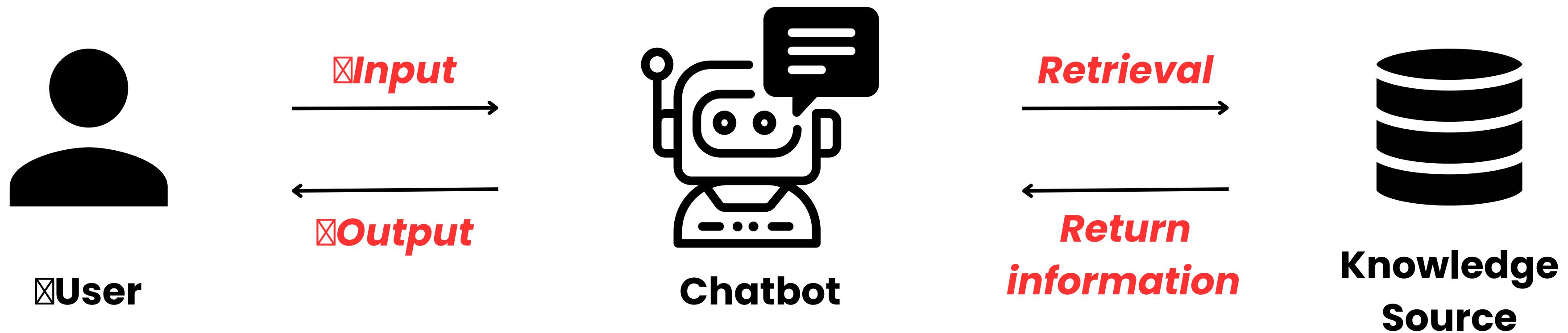
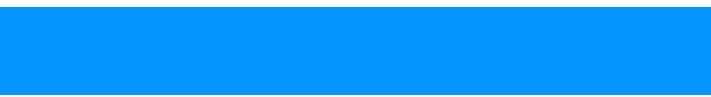
 **Back-End**



 **Front-End**

# MODEL

RECAP: RAG



## Retrieval Augmented Generation

# MODEL

## RECAP: RAG PROGRESS

🤖 Chat-pub Chatbot Demo



안녕?



안녕하세요! 저는 정책 혜택에 대해 안내해주는 챗봇입니다. 궁금한 점이 있으시면 언제든지 물어보세요. 저는 친절하게 답변해드릴게요!



난 지금 고양시에 살고 있고 20대 남성이다. 내가 받을 수 있는 정책 혜택이 있을까?



고양시에 거주하고 있는 20대 남성이신 경우, '고양청년 창업 재정지원 프로그램'에 참여하여 혜택을 받을 수 있습니다. 이 프로그램은 창업을 준비하는 청년들에게 무담보 신용보증과 최대 5천만 원의 대출을 지원 합니다. 자세한 신청 방법은 경기신용보증재단 고양지점을 방문하여 상담하시면 됩니다. 자세한 내용은 [고양시 공식 홈페이지](#)에서 확인하실 수 있습니다.



제목: 고양청년 창업 재정지원 프로그램

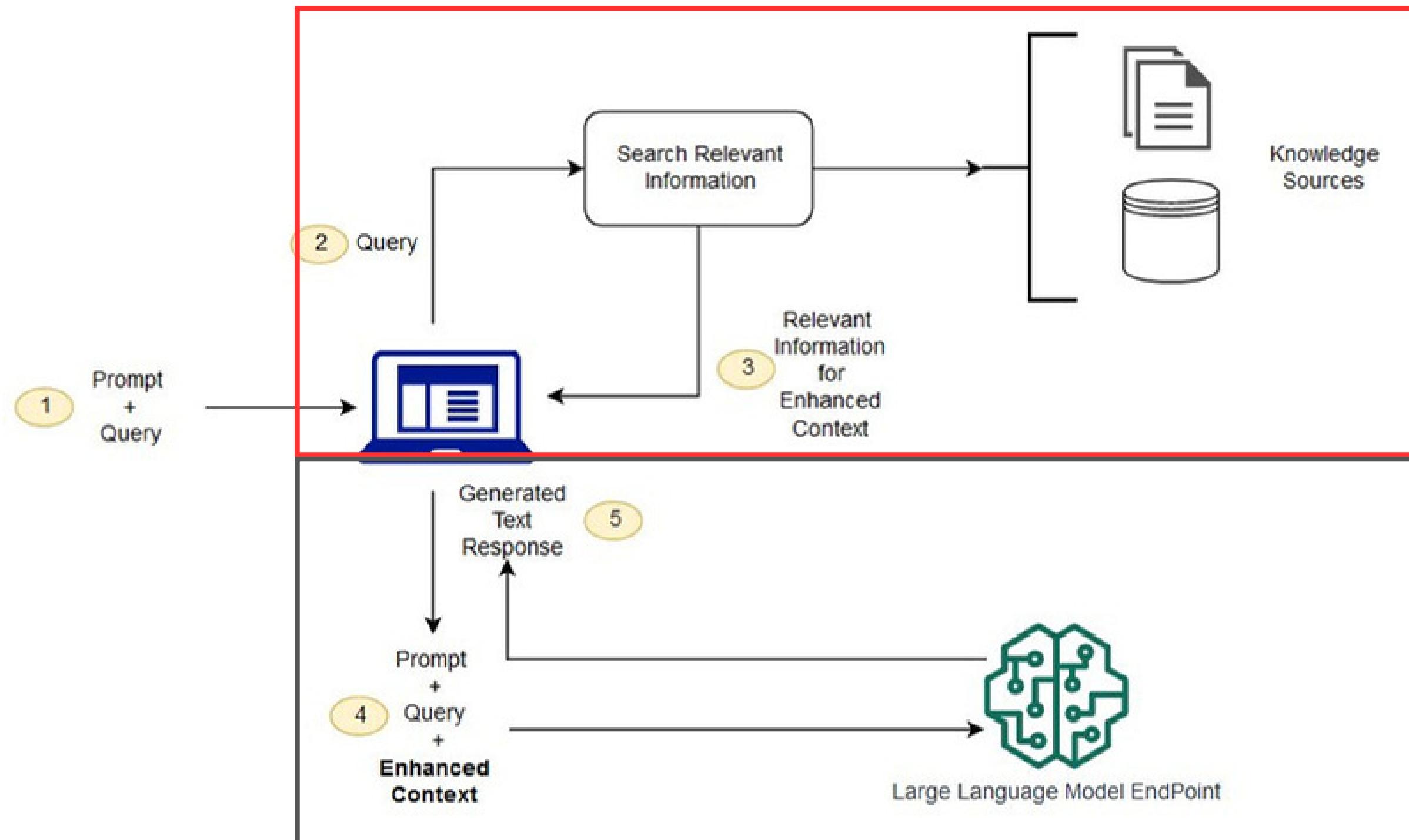
링크:<https://www.youthcenter.go.kr/youngPlcyUnif/youngPlcyUnifDtl.do?pageIndex=1&bizId=R2023060213115>

난 지금 고양시에 살고 있고 20대 남성이다. 내가 받을 수 있는 정책 혜택이 있을까?

☒Demo

# MODEL

## RECAP: RAG



**[Retrieval]**

**Sentence  
Transformers**

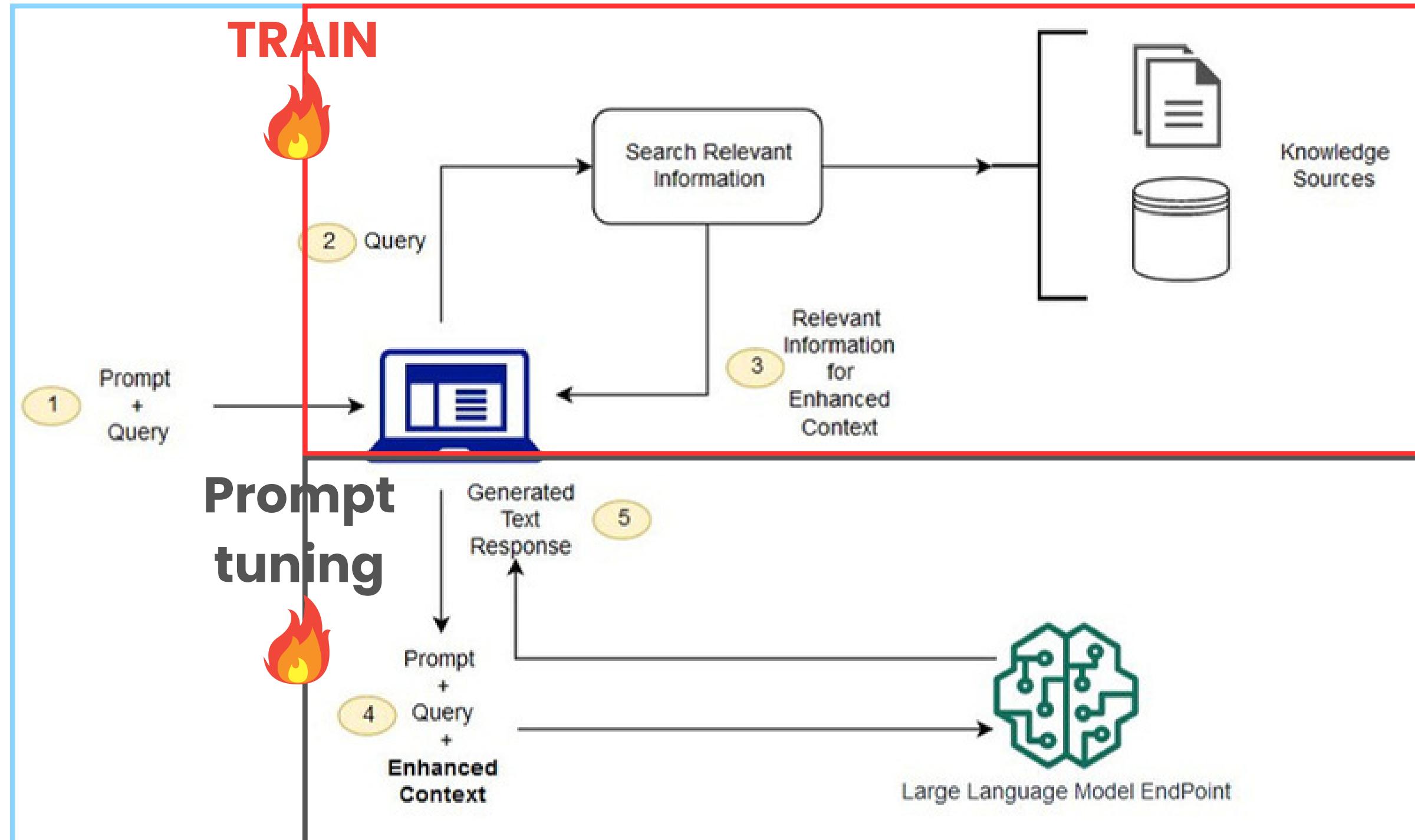
**[Generation]**

**Open AI API  
(GPT-4)**

# MODEL

## RAG TRAINING STRATEGY

TEST



[Retrieval]

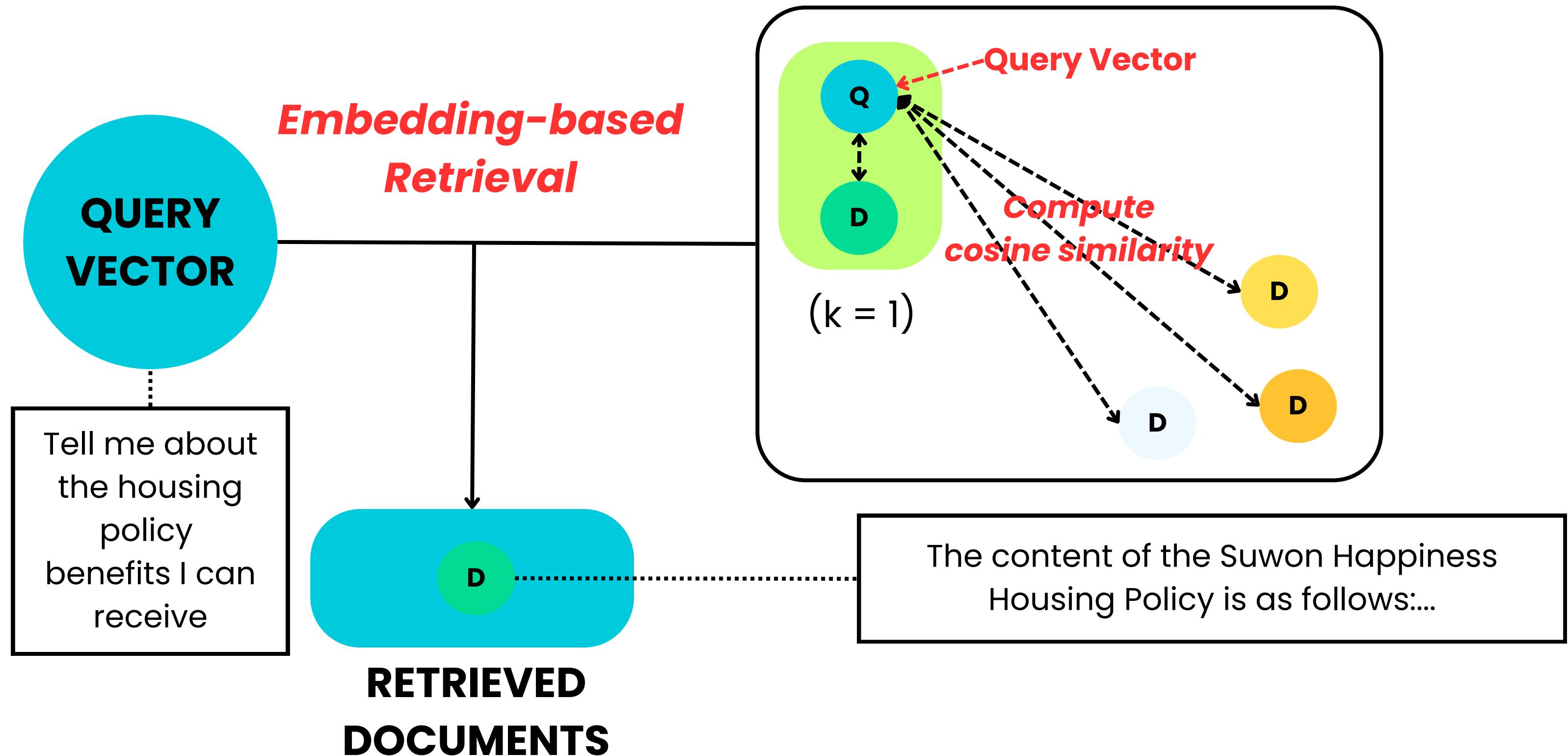
Sentence  
Transformers

[Generation]

Open AI API  
(GPT-4)

# MODEL

## RECAP: RETRIEVAL



# MODEL

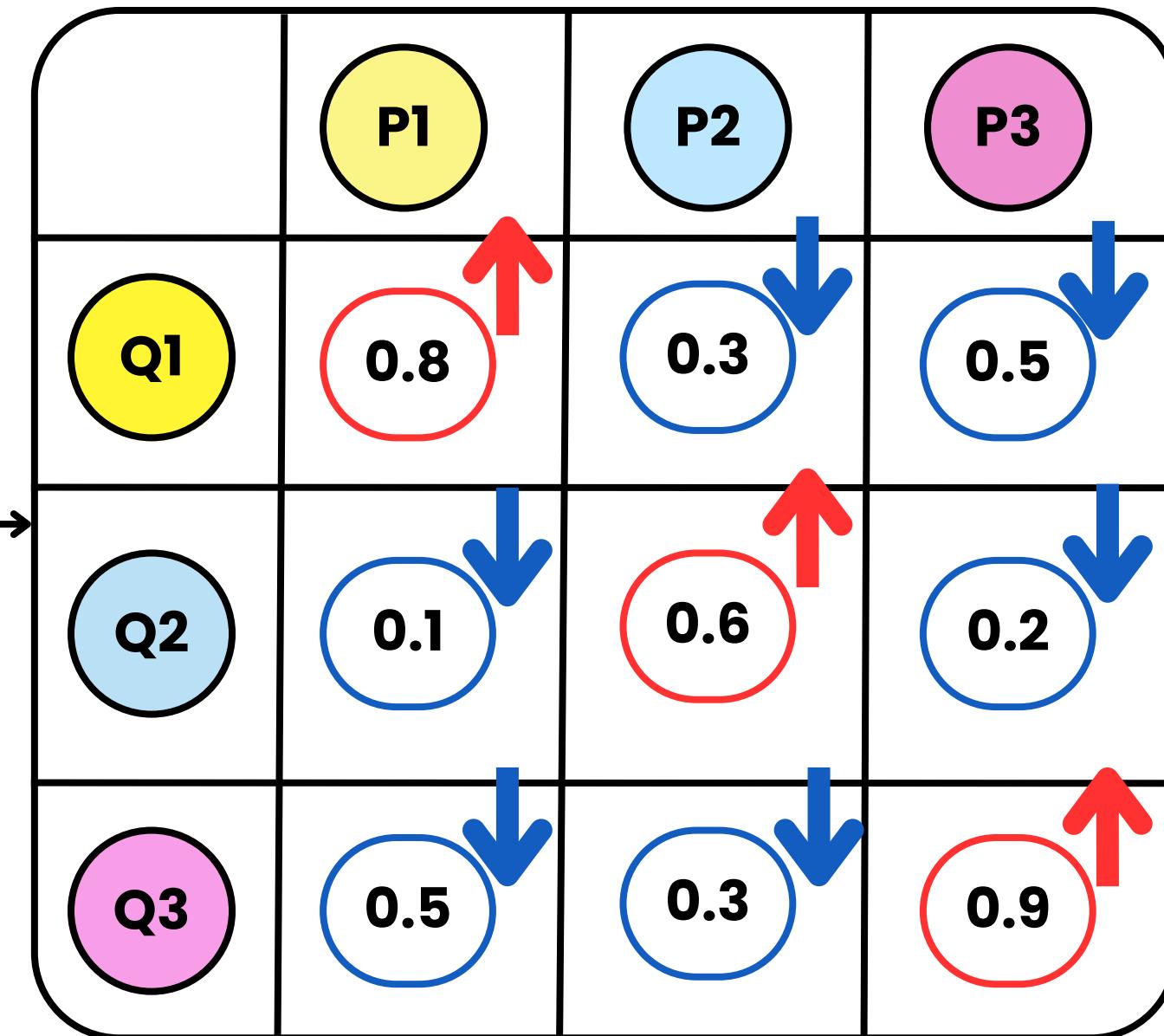
## RECAP: SENTENCE TRANSFORMERS

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

## COSINE SIM

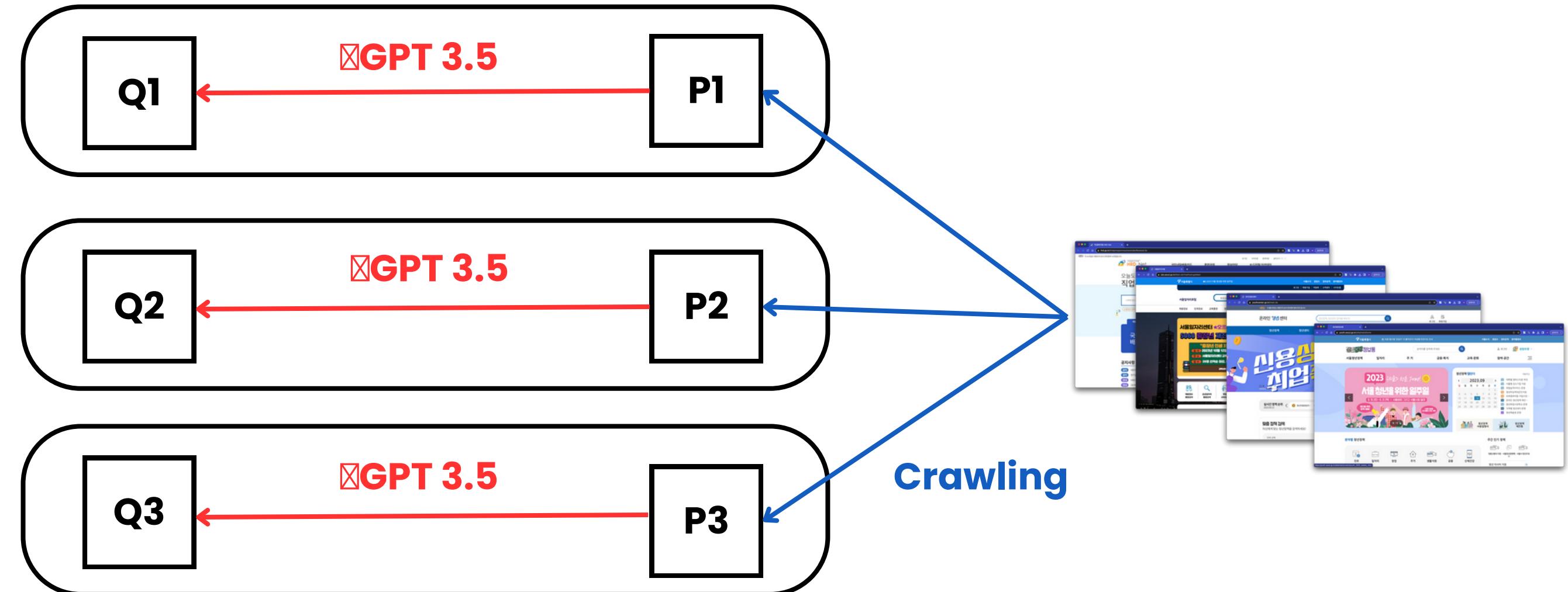


SENTENCE  
TRANSFORMERS



# MODEL

TRAIN: SENTENCE TRANSFORMERS



## EXAMPLE

**Q1:** What's Sungkyunkwan University?

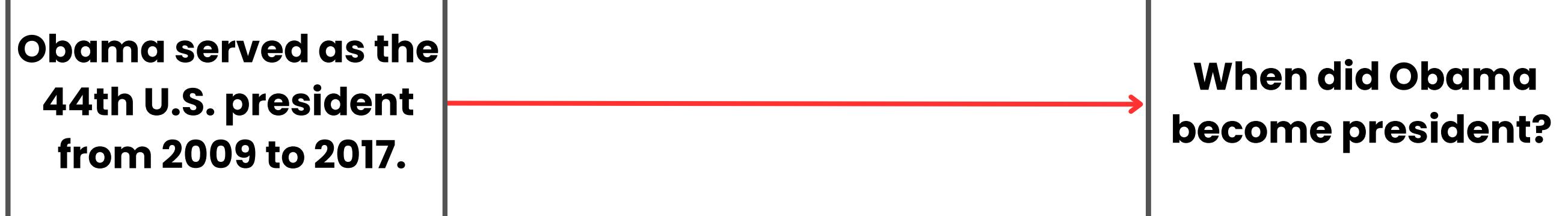
**P1:** Sungkyunkwan University is one of the best universities in Korea.

# MODEL

TRAIN: SENTENCE TRANSFORMERS

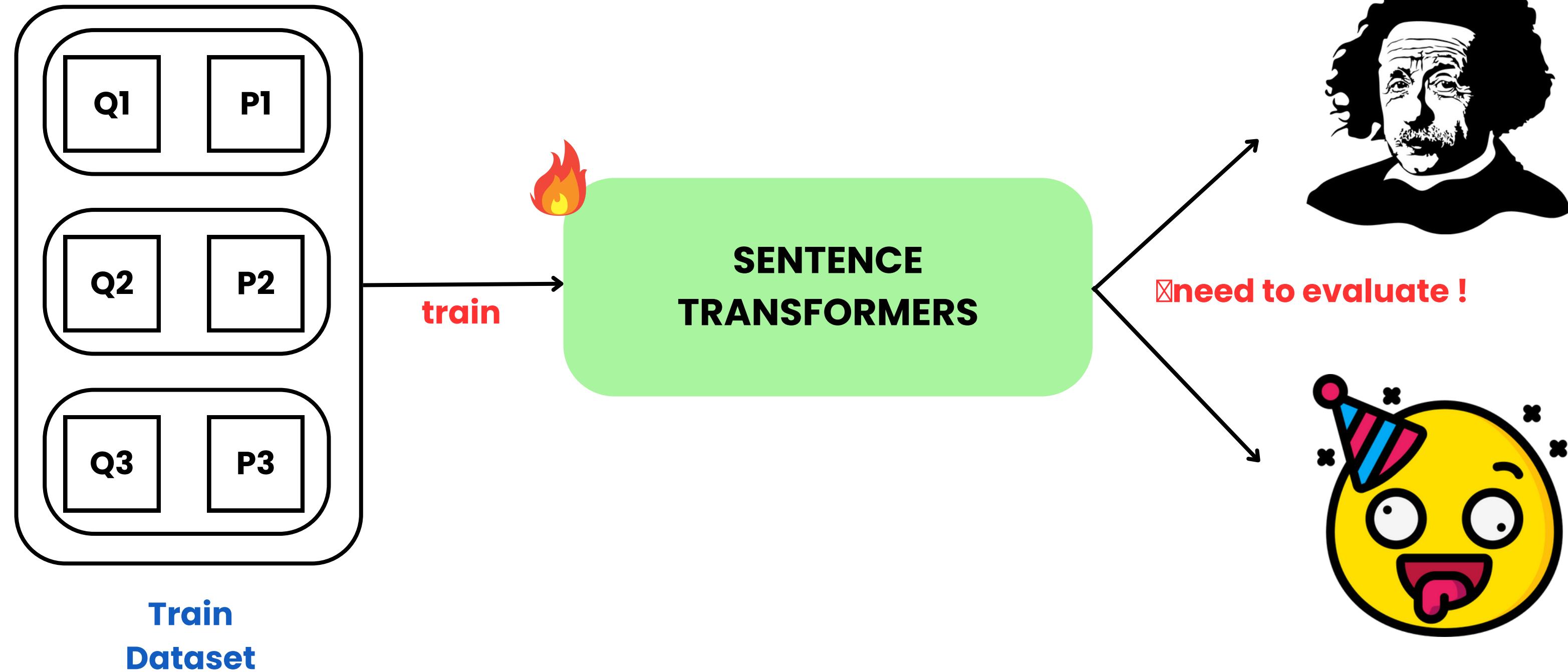


## EXAMPLE



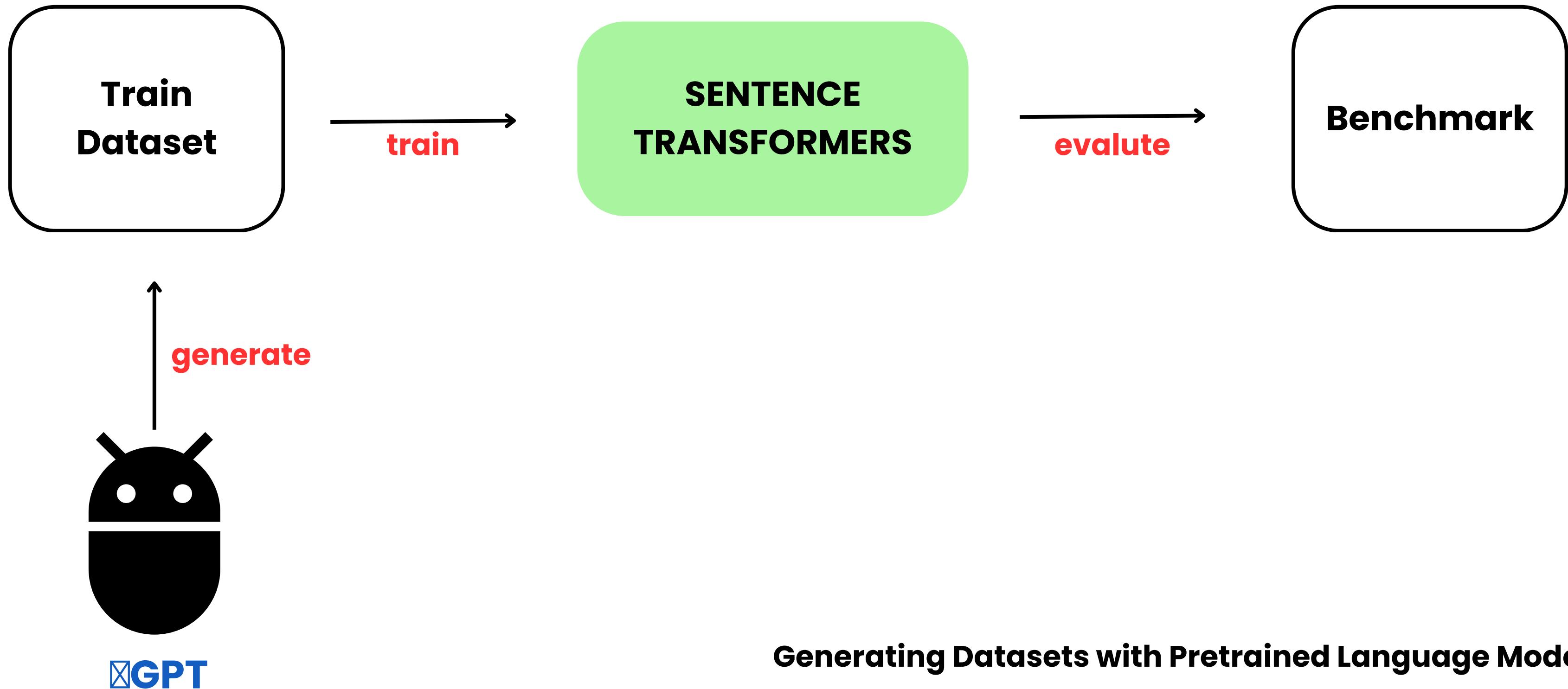
# MODEL

TRAIN: SENTENCE TRANSFORMERS



# MODEL

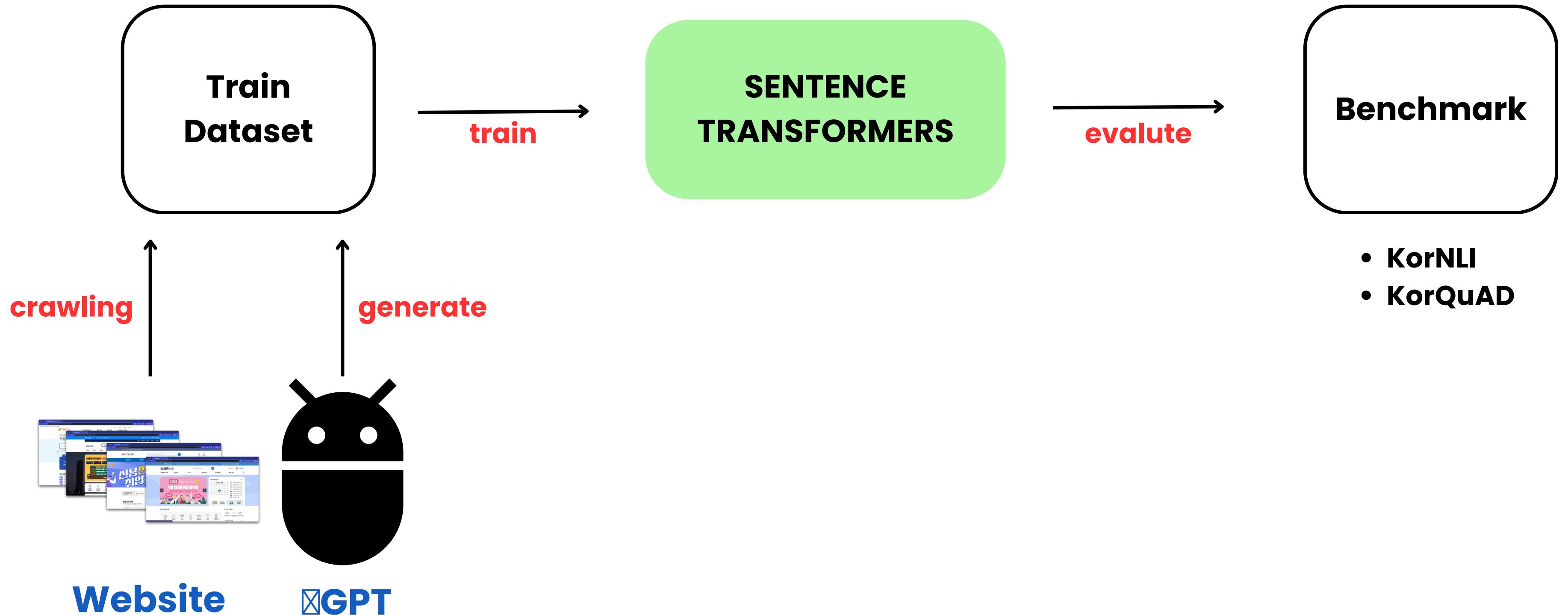
TRAIN: SENTENCE TRANSFORMERS



Generating Datasets with Pretrained Language Models

# MODEL

TRAIN: SENTENCE TRANSFORMERS



# MODEL

PROMPT TUNING: LLM

Answer the user's questions using the given context.  
If you do not know the answer reply with 'I am sorry'.

Context:

Retrieved  
Snippets

Question:

User  
question



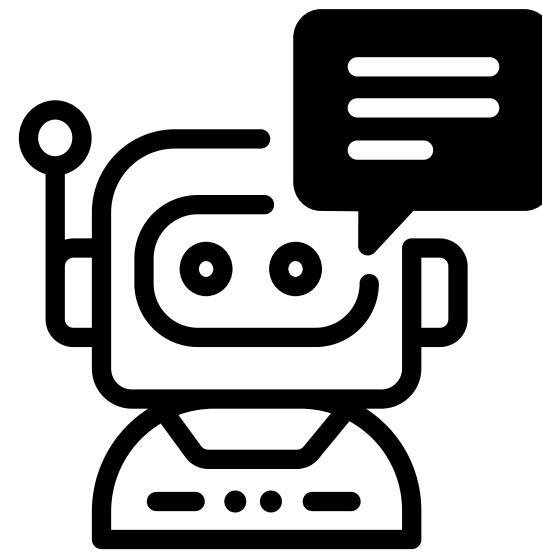
Baselin

e

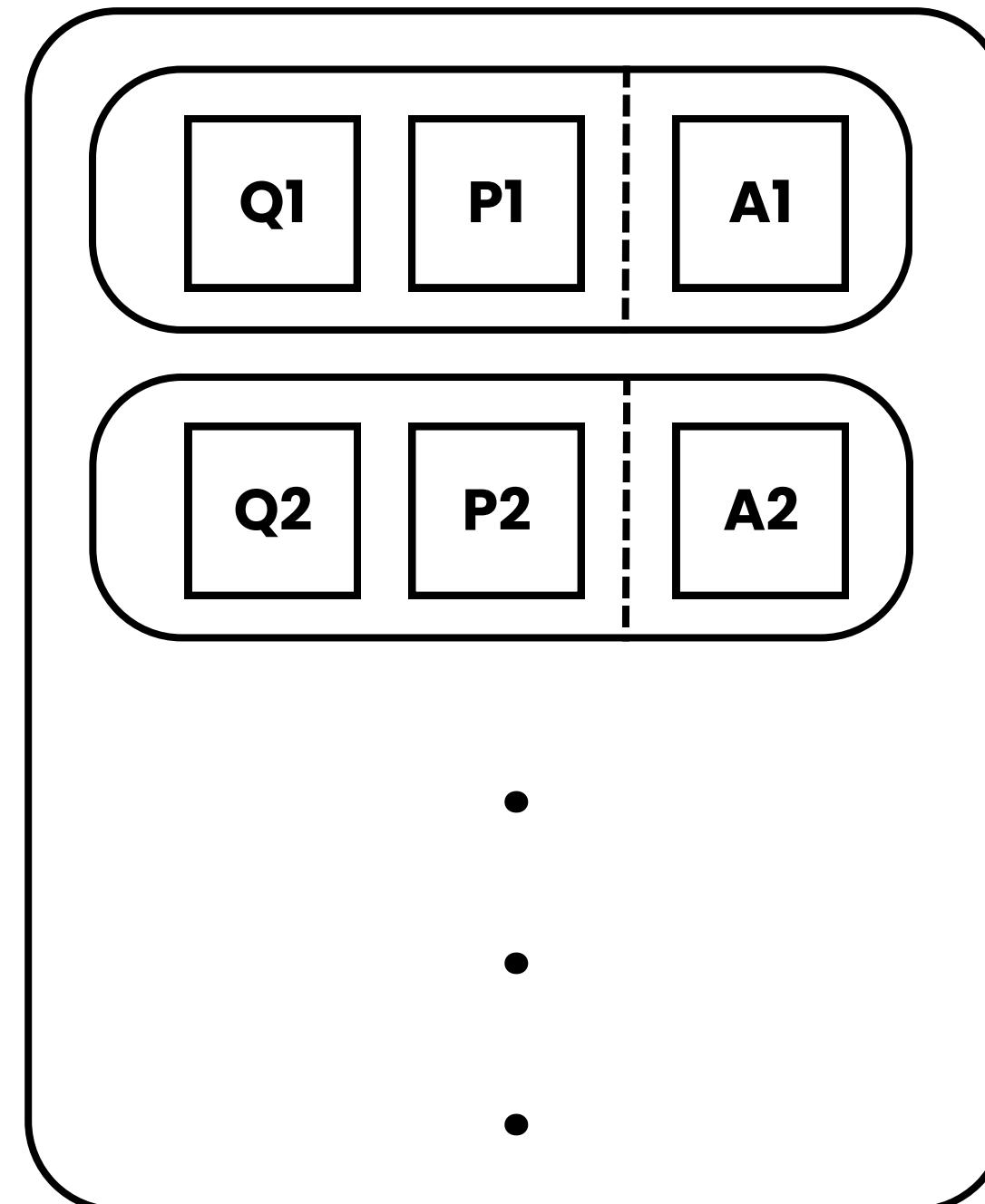
[HTTPS://WWW.LANGCHAIN.COM/](https://www.langchain.com/)

# MODEL

## PROMPT TUNING: LLM



LLM



## EXAMPLE

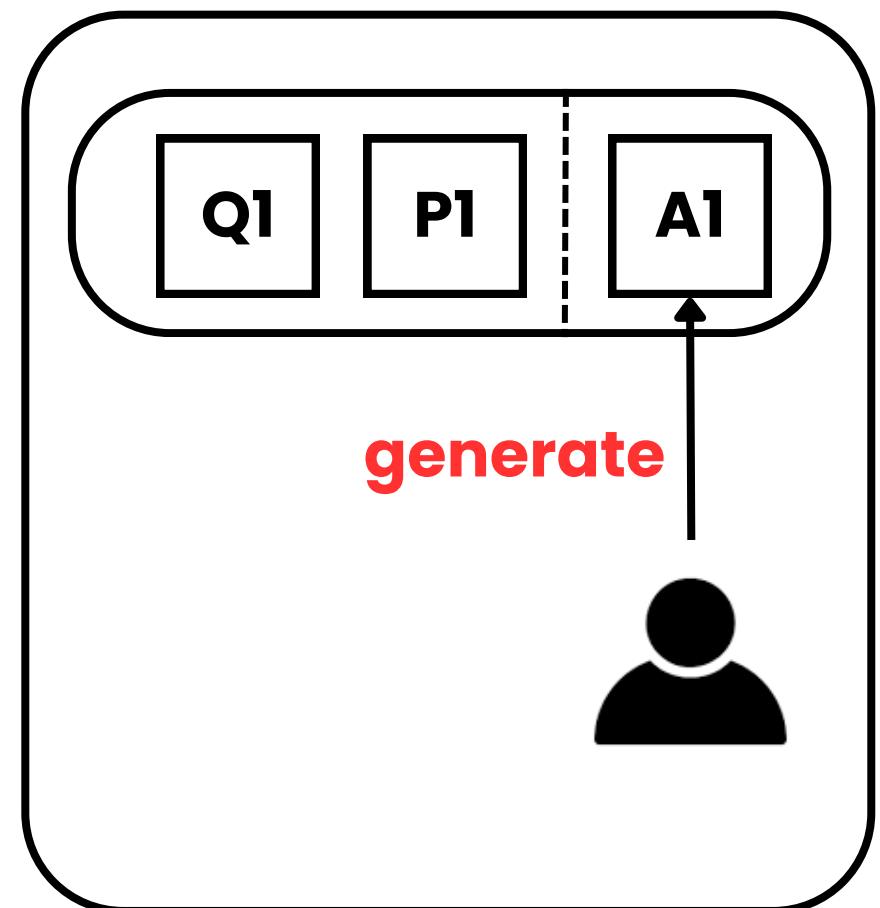
**Q1:** What's Sungkyunkwan University?

**P1:** Sungkyunkwan University is one of the best universities in Korea.

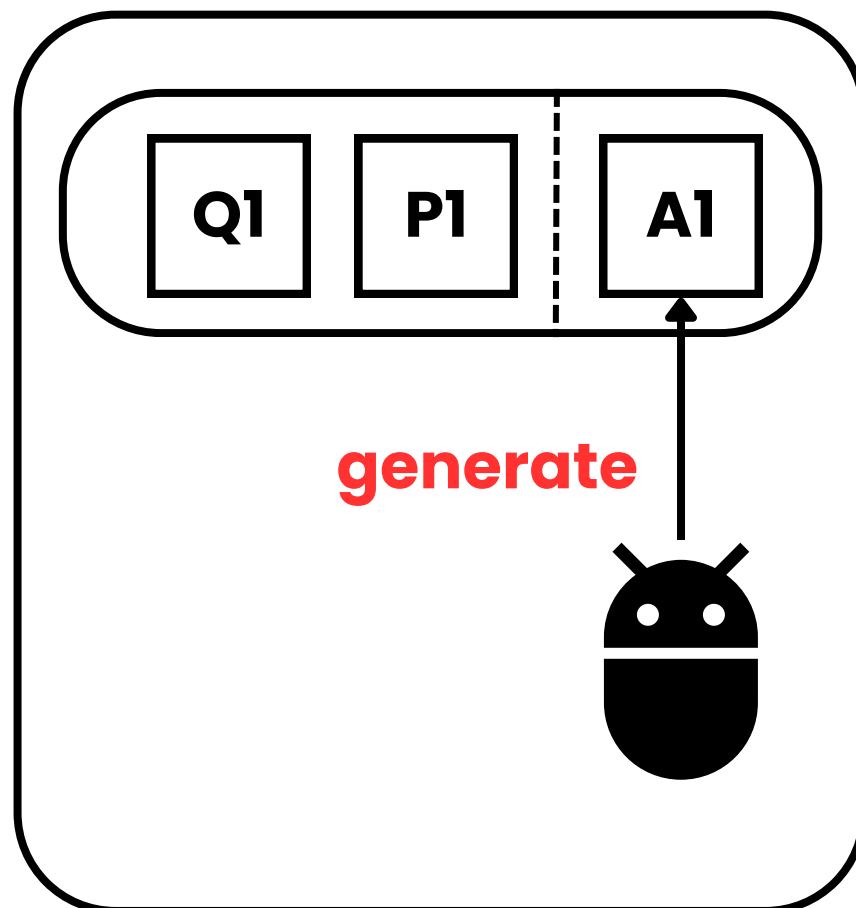
**A1:** One of the best Univ in Korea.

# MODEL

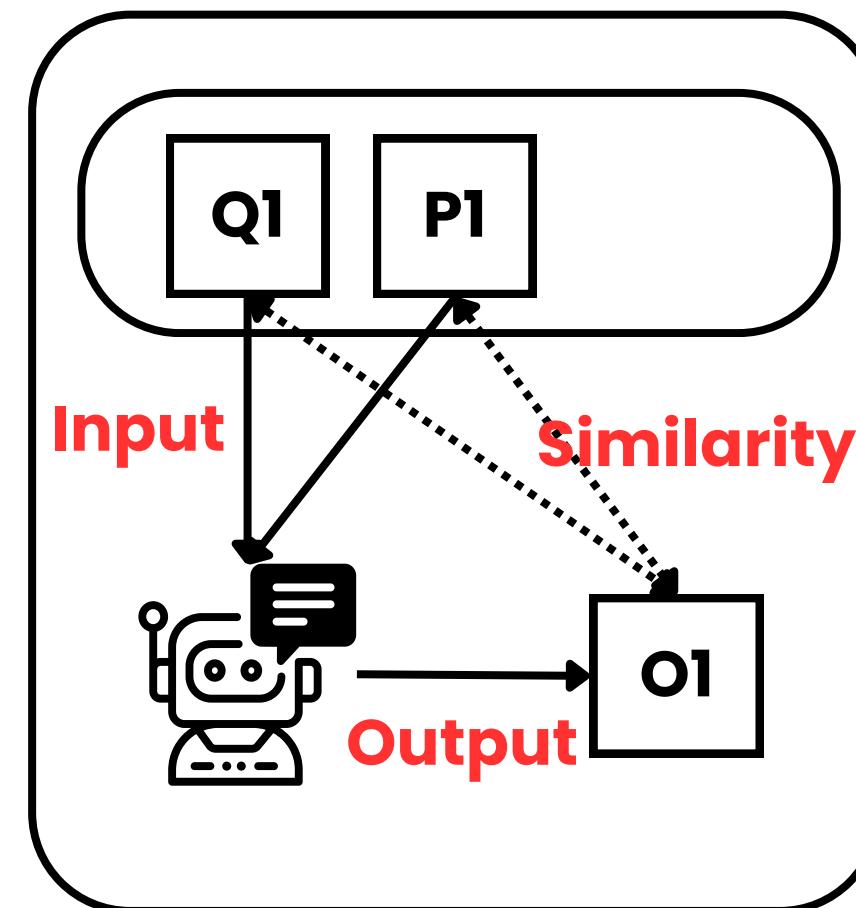
PROMPT TUNING: LLM



Idea 1

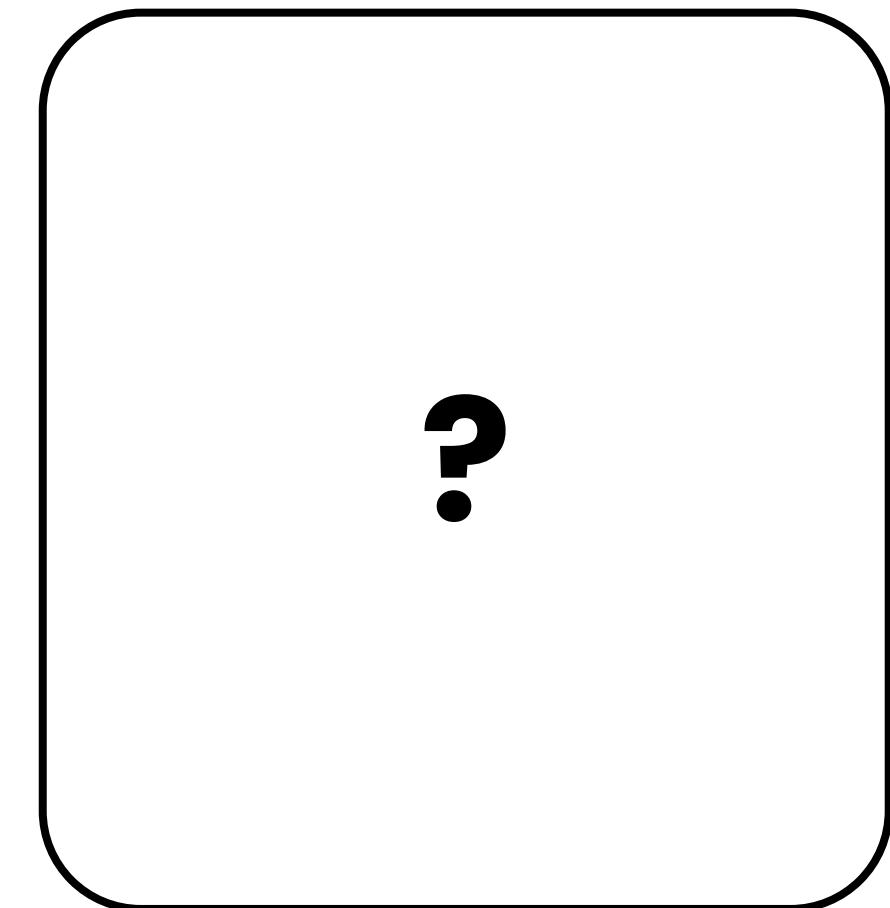


Idea 2



Idea 3

...



Idea N

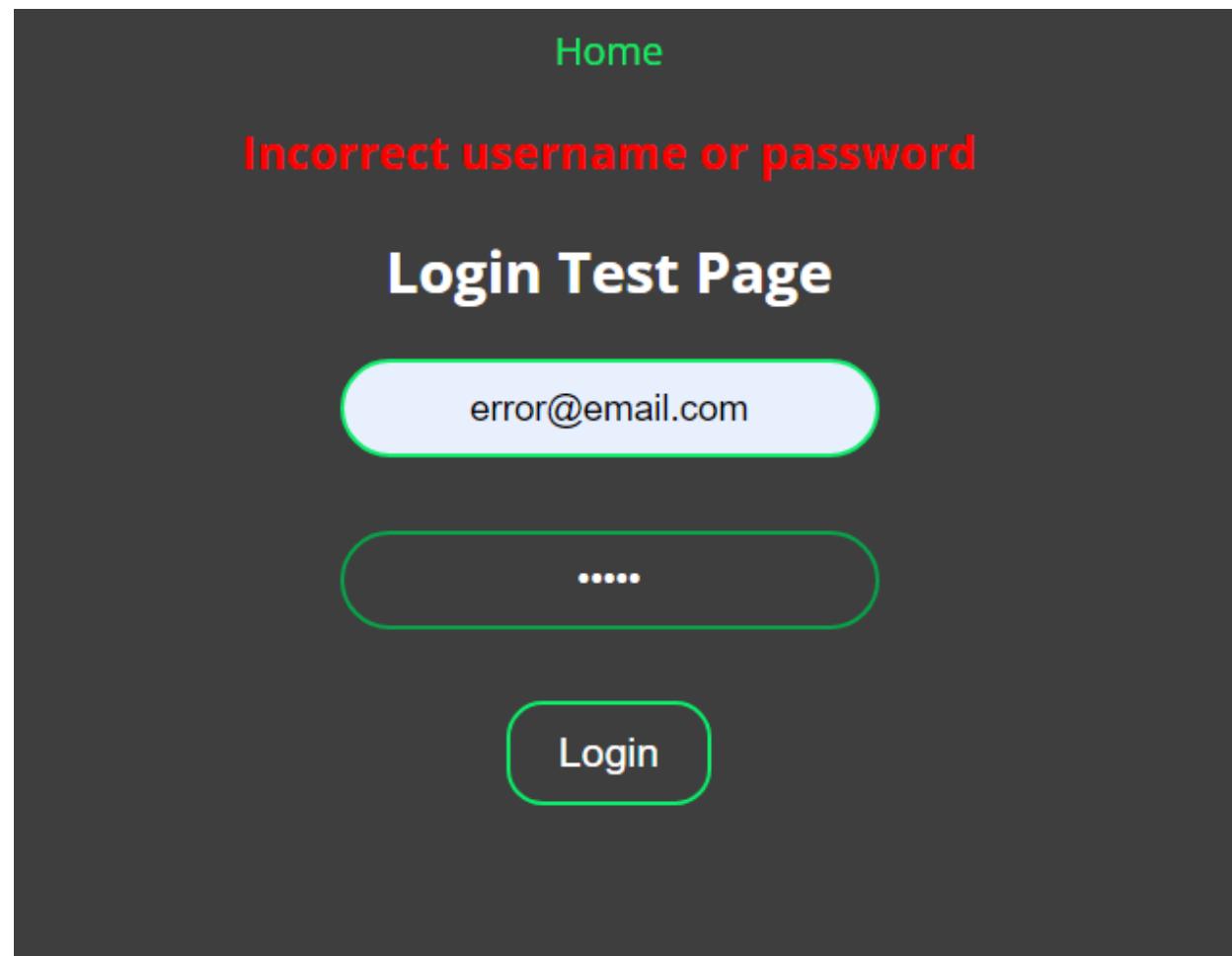
# **BACKEND COMPONENT**

- **LOG IN, SIGN-UP**
- **HASH PASSWORD AND LINK MARIADB**
- **DB STRUCTURE**

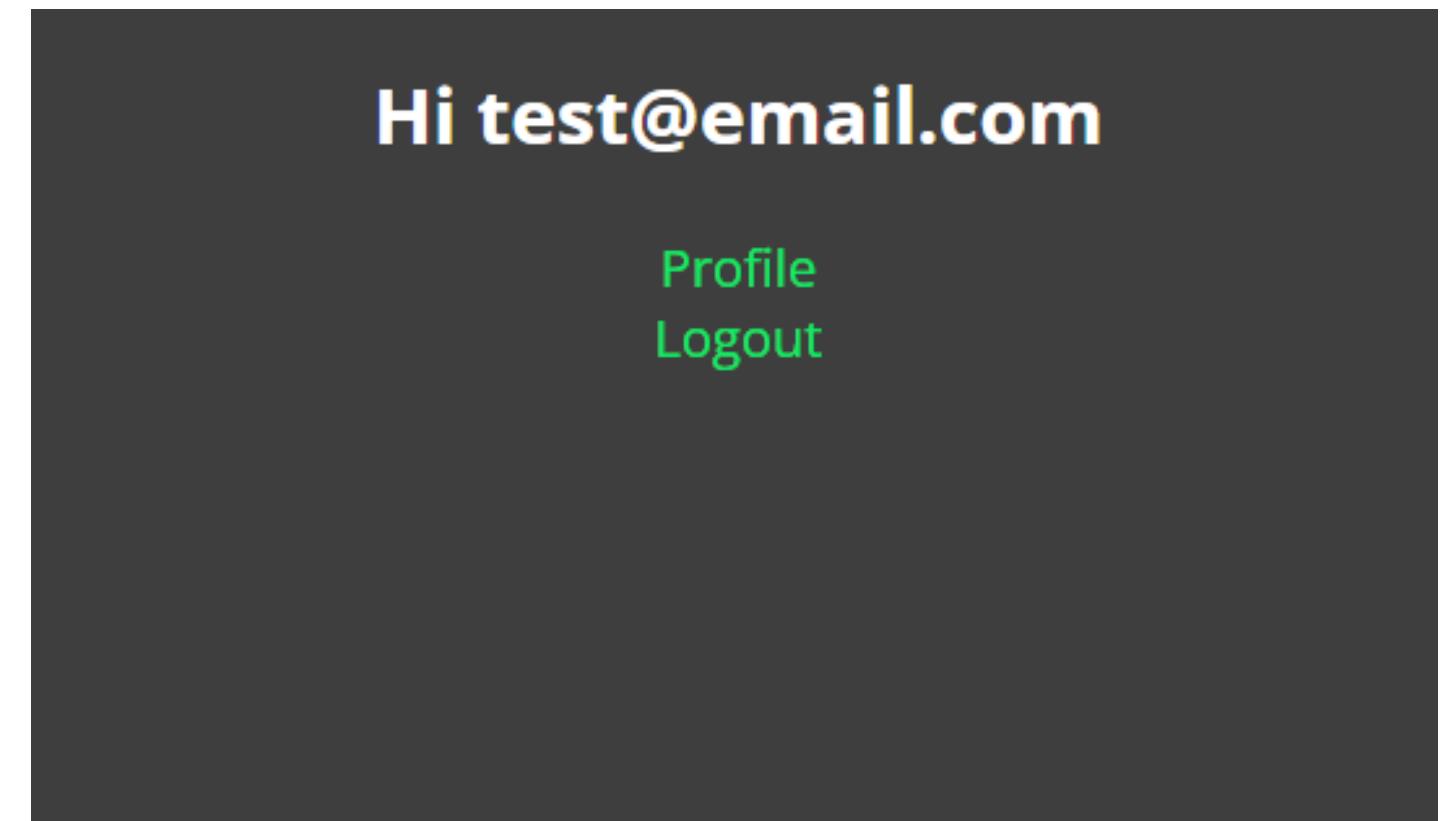
# BACKEND

## LOGIN

### NO & WRONG ACCOUNT



### SUCCESS LOGIN



# BACKEND SIGN-UP

## DUPLICATE ACCOUNT

Home

A user already has this username

Sign up Test Page

test@email.com

.....

.....

Signup

This screenshot shows a sign-up form with a dark background. At the top right is a blue bar. The page title 'Sign up Test Page' is at the top. Below it is a red error message 'A user already has this username'. There are three input fields: an email field containing 'test@email.com', a password field with four dots, and a confirmation field with four dots. A large green 'Signup' button is at the bottom.

## NOT MATCH EMAIL FORM

Home

The Email don't match the form

Sign up Test Page

12345

.....

.....

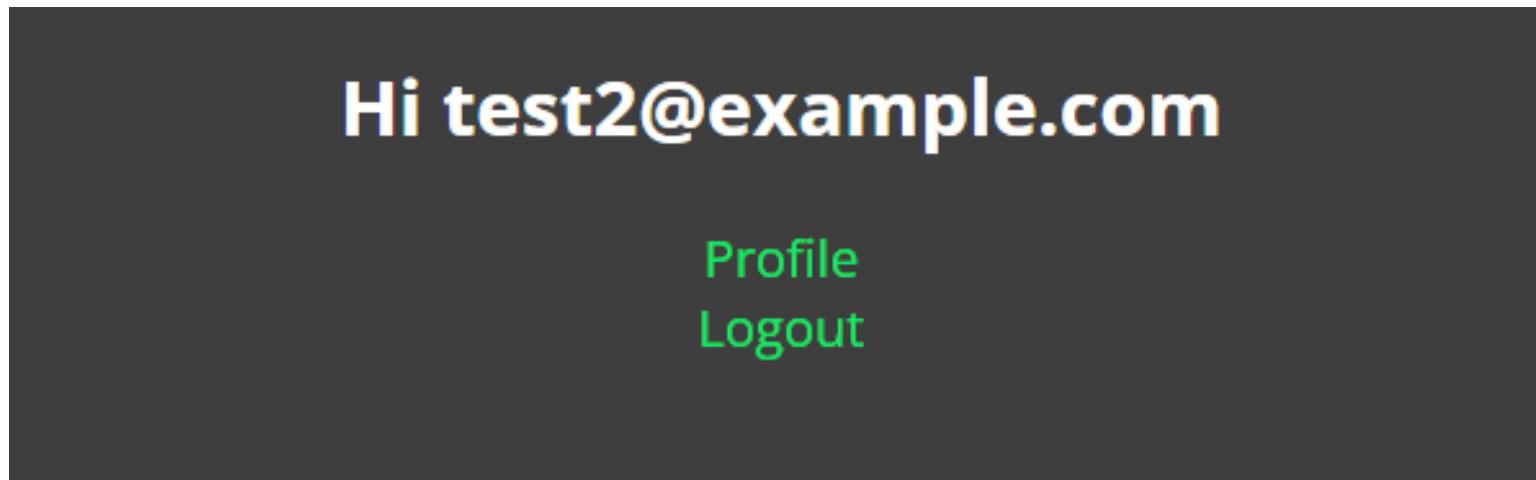
Signup

This screenshot shows a sign-up form with a dark background. At the top right is a blue bar. The page title 'Sign up Test Page' is at the top. Below it is a red error message 'The Email don't match the form'. There are three input fields: a password field containing '12345', a password field with four dots, and a confirmation field with four dots. A large green 'Signup' button is at the bottom.

# BACKEND

## SIGN-UP AND HASH PASSWORD

### SUCCESS SIGN-UP



```
MariaDB [test]> select * from users;
+-----+-----+
| user_id | Username          | password           |
+-----+-----+
|      11 | test@email.com    | 5994471abb01112afcc18159f6cc74b4f511b99806da59b3caf5a9c173cacfc5 |
|      12 | test2@example.com | 5994471abb01112afcc18159f6cc74b4f511b99806da59b3caf5a9c173cacfc5 |
+-----+-----+
2 rows in set (0.001 sec)
```

# BACKEND DATABASE STRUCTURE OF POLICY DATA

The screenshot shows the database structure for the ChatPub database. It includes:

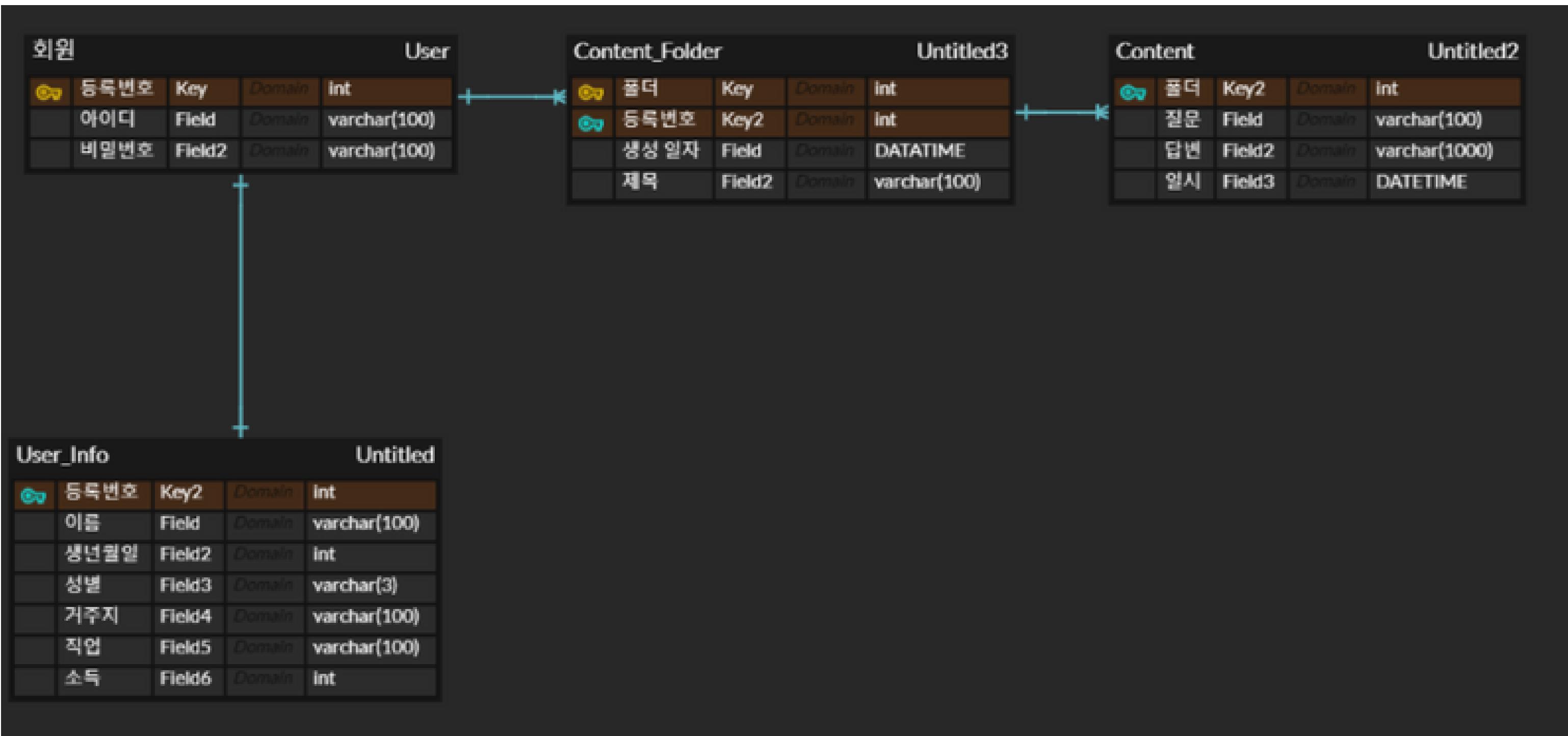
- Databases: ChatPub
- Tables:
  - yp\_all\_overview (496K)
  - yp\_etc (320K)
  - yp\_methods (464K)
  - yp\_qualification (1.5M)
  - yp\_summary (496K)
- Views
- Indexes

The screenshot shows the data grid for the yp\_all\_overview table in the ChatPub database. The table has columns: Grid, yp, title, and r\_num.

Grid	yp	title	r_num
1	0	'청춘마이크' 사업을 통한 청년문화예술가 지원	R2023012
2	1	(가칭)금천청년꿈터 조성	R2023050
3	2	(건강보험) 임의계속가입제도	R2023012
4	3	(광진구)미취업 청년 어학시험 응시료 지원사업	R2023050
5	4	(국민연금) 실업크레딧 지원	R2023030
6	5	(동·하계)대학생 행정체험 연수 운영	R2023060
7	6	(서울 관악구)지역주도형 청년일자리사업	R2023050
8	7	(서울 관악구)청년네트워크 운영	R2023050
9	8	(서울 관악구)청년정책위원회 운영	R2023050

# BACKEND

## DATABASE STRUCTURE OF USER



# **BACKEND CHALLENGE&SOLUTION**

- **UNDERSTANDING NEXTJS -> USING TEXTBOOK**
- **NEED FOR PYTHON SERVER -> NEXTJS + FASTAPI**

# FRONT-END MAIN PAGE

Welcome, John Doe!

CHAT - PUB

Ver 1.0 Oct 120

**Examples**  
What are some policies that will benefit me?

**Capabilities**  
Remembers what user said earlier in the conversation

**Limitations**  
May occasionally generate incorrect information

How can I apply for housing aid for Seoul residents?

Allows user to provide follow-up corrections

It does not guarantee. Please use with caution.

Can you tell me about any job education that is available for me?

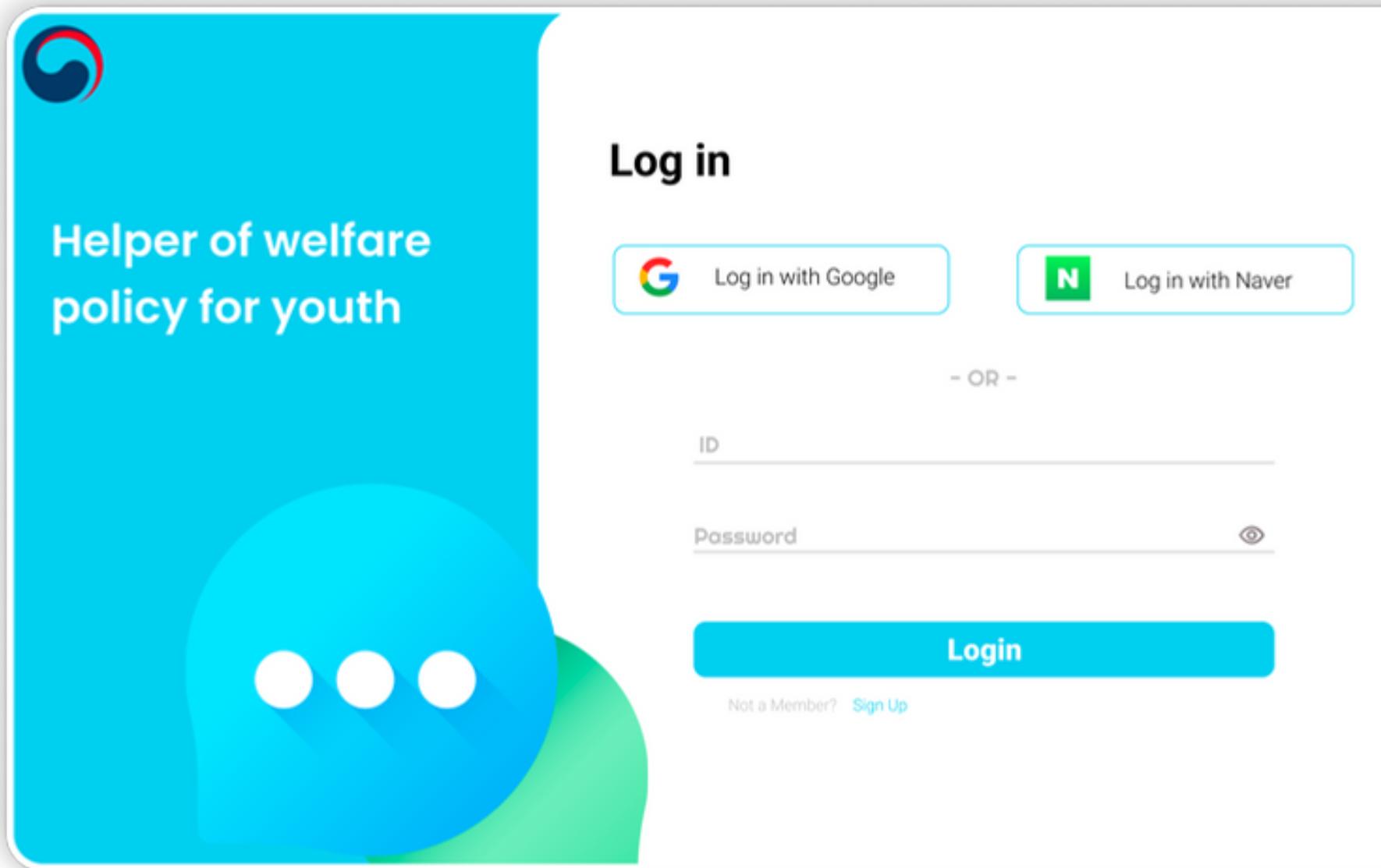
Trained to decline inappropriate requests

Limited knowledge of policy

New dialog

Send a message

# FRONT-END LOGIN PAGE / SIGN UP PAGE



The login page features a teal header with the text "Helper of welfare policy for youth". Below the header is a large blue circular graphic containing three white dots. The main form area has a light gray background. It includes a "Log in" heading, two social login buttons ("Log in with Google" and "Log in with Naver"), a "Full Name" input field, a "Email Address" input field, a "Password" input field with an "Eye" icon, and a "Login" button. A "Not a Member? Sign Up" link is located at the bottom left.

**Log in**

 Log in with Google    Log in with Naver

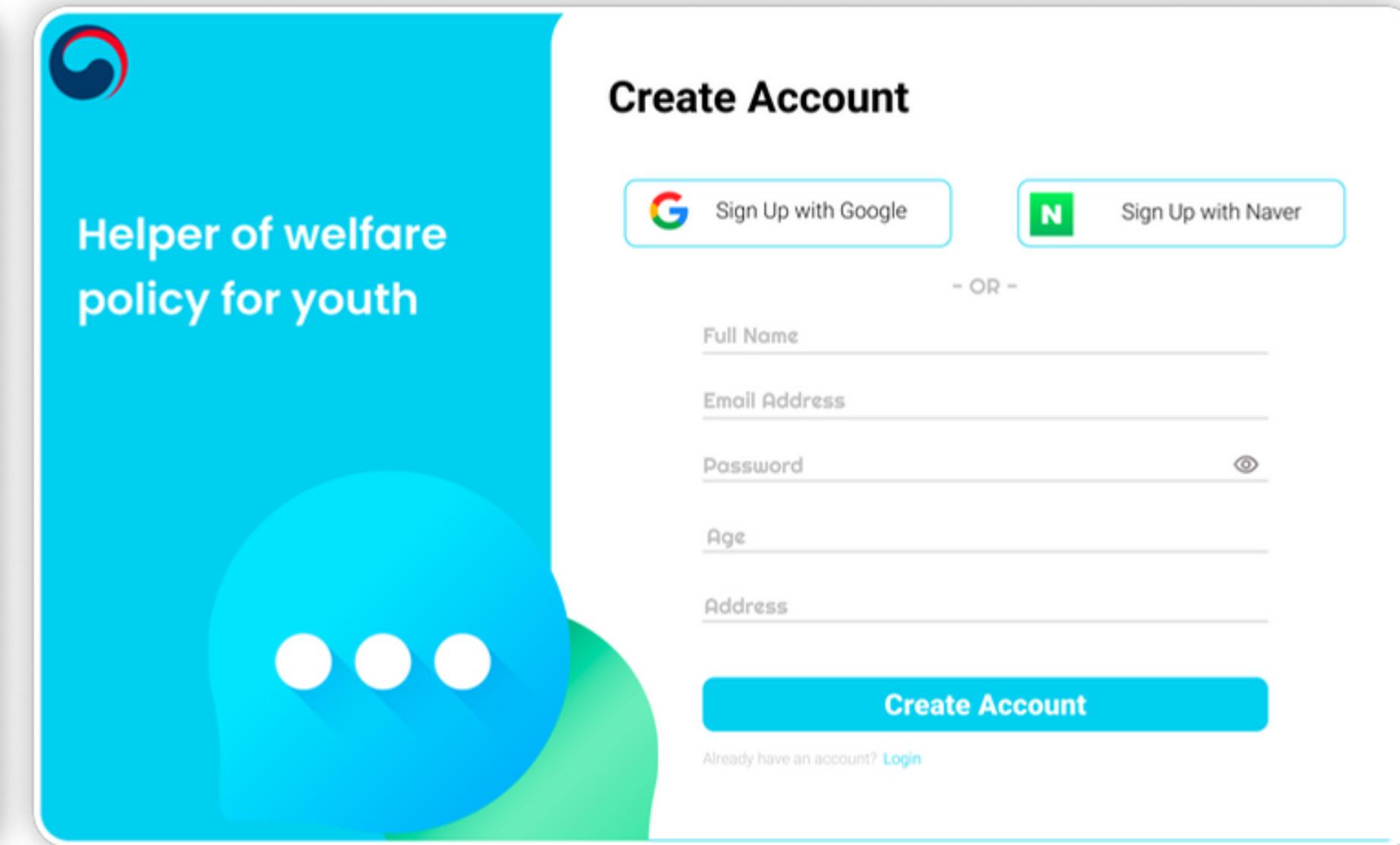
- OR -

ID

Password  

**Login**

Not a Member? [Sign Up](#)



The sign up page features a teal header with the text "Helper of welfare policy for youth". Below the header is a large blue circular graphic containing three white dots. The main form area has a light gray background. It includes a "Create Account" heading, two social sign up buttons ("Sign Up with Google" and "Sign Up with Naver"), a "Full Name" input field, an "Email Address" input field, a "Password" input field with an "Eye" icon, an "Age" input field, an "Address" input field, and a "Create Account" button. A "Already have an account? Login" link is located at the bottom left.

**Create Account**

 Sign Up with Google    Sign Up with Naver

- OR -

Full Name

Email Address

Password  

Age

Address

**Create Account**

Already have an account? [Login](#)

# FRONT-END SETTING PAGE

The screenshot shows a user interface for managing personal information. At the top right, a blue bar displays the welcome message "Welcome, John Doe!". On the left, a vertical sidebar contains four icons: a house (Account), three dots (Personal info), a magnifying glass (Search), and a gear (Settings). The main content area is titled "Personal info". It includes sections for "Legal name" (John Doe, Edit), "Email address" (JohnDoe@gmail.com, Edit), "Phone numbers" (010-XXXX-XXXX, Add), "Address" (Not provided, Edit), and "Emergency contact" (Not provided, Add).

Welcome, John Doe!

Account > Personal info

## Personal info

Legal name [Edit](#)  
John Doe

Email address [Edit](#)  
JohnDoe@gmail.com

Phone numbers [Add](#)  
010-XXXX-XXXX

Address [Edit](#)  
Not provided

Emergency contact [Add](#)  
Not provided

# FRONT-END MESSAGE PAGE

The screenshot displays a messaging interface with a blue header bar containing the text "Welcome, John Doe!". Below the header, there is a list of messages. The messages are color-coded: light gray for the user and teal for the bot. Each message includes a timestamp and a small profile icon.

**User Message:** Hi, I'm a 25 and I currently live in Gyeonggi-do.  
2023-11-02

**Bot Response:** 10:00 AM, 20 Oct  
Hi, I'm a 25 and I currently live in Gyeonggi-do.  
I wonder if there's a housing policy I can benefit from?  

**User Message:** Could you tell me the exact place of residence?  
10:01 AM, 20 Oct

**Bot Response:** 10:01 AM, 20 Oct  
I live in Suwon  

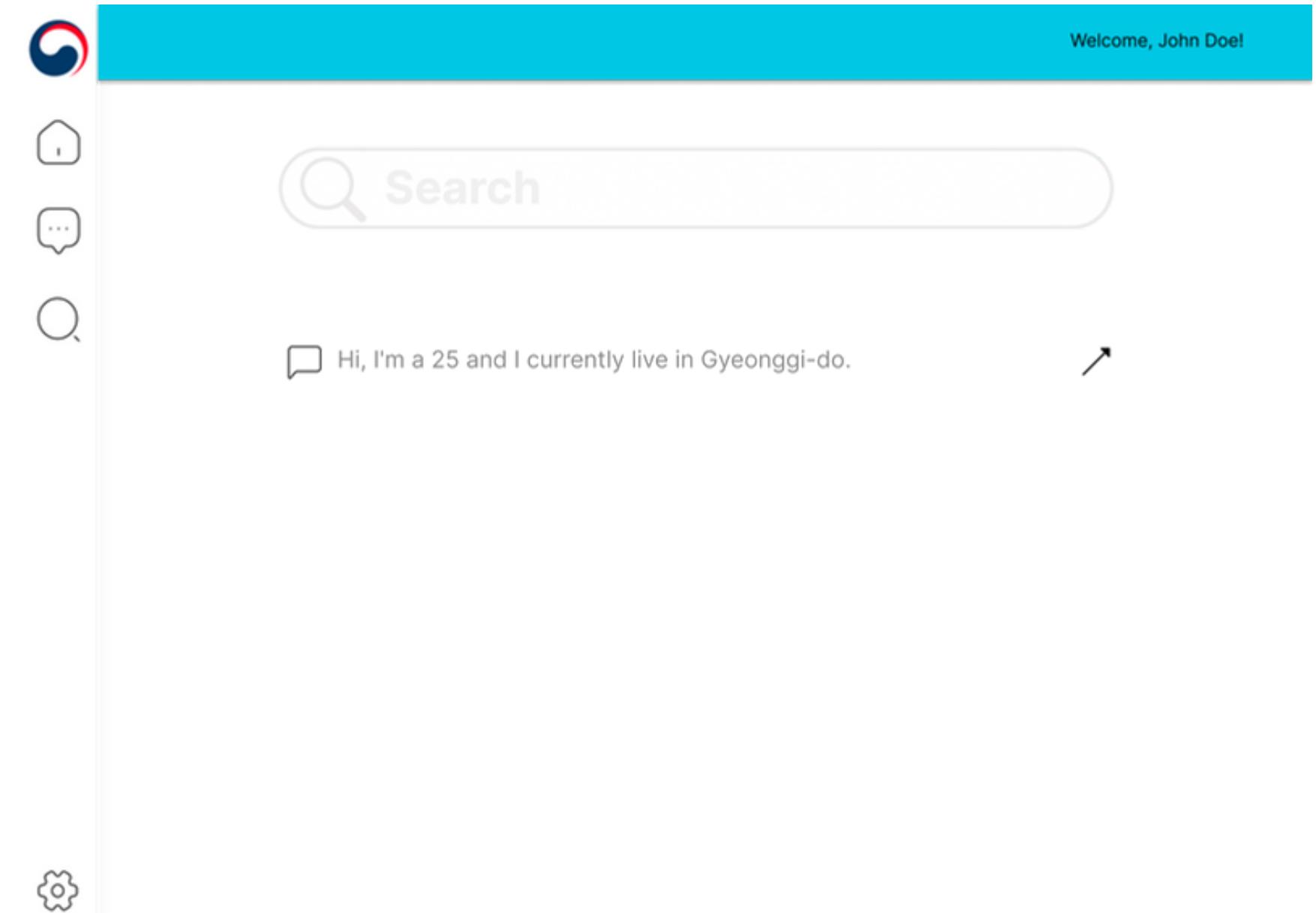
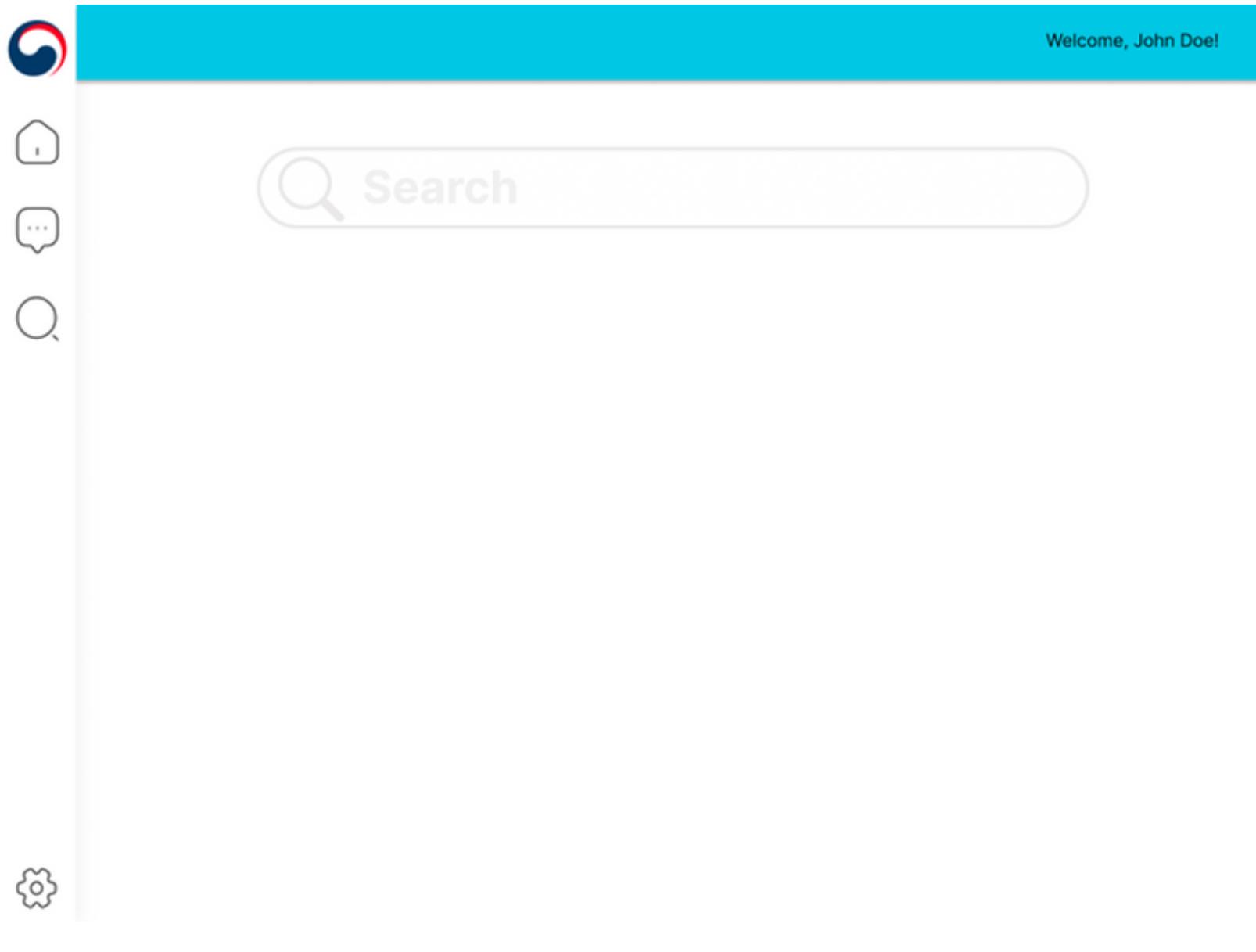
**Bot Response:** 10:01 AM, 20 Oct  
Here is Policy that might benefit you!  
the policy is called recruitment of youth zones that is for To promote housing stability of young people in Suwon City to 50% level of young people in Suwon City.

**Bottom Bar:**

- Generate message..
- Description
- ESC or Click to cancel



# FRONT-END SEARCH PAGE



# FRONT-END CHALLENGE&SOLUTION



**Monotony**: Working with a single color can risk making the design appear monotonous or boring.



**MAINTAINING VISUAL INTEREST WHILE USING LESS COLOR**



**Texture and Depth**: Creating texture and a sense of depth in a one-colored design can be challenging.



**RELY ON TECHNIQUES LIKE SHADING AND PATTERNS**



**Clarity and Readability**: Achieving clarity and readability in a one-colored design, especially with text, can be tough.



**BALANCE THE COLOR CHOICE WITH THE CONTENT TO ENSURE IT'S EASILY LEGIBLE.**



# LIMITATION



## Resource Constraints



**Limitations on the information due to fetching data from a single web page**



## Privacy and Security



**Prone to malicious attacks**



## The difficulty of model training



**Not being able to train on real user data**

# PROJECT SCHEDULE



# QUESTIONS & ANSWERS



# APPENDIX

- **Automatically augmenting the reference question with large language models is surprisingly effective, showing higher correlations with human annotated scores.**

Shinhyeok Oh, Hyojun Go, Hyeongdon Moon, Yunsung Lee, Myeongho Jeong, Hyun Seung Lee, and Seungtaek Choi. 2023. *Evaluation of Question Generation Needs More References*. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 6358–6367, Toronto, Canada. Association for Computational Linguistics.

- **A ChatGPT-based QG model achieves state-of-the-art (SOTA) performance in generating answerable questions**

<https://arxiv.org/pdf/2309.12546v1.pdf>

# APPENDIX

**Task:** Write two sentences that mean the same thing.

**Sentence 1:** “A man is playing a flute.”

**Sentence 2:** “He’s playing a flute.”

**Task:** Write two sentences that are somewhat similar.

**Sentence 1:** “A man is playing a flute.”

**Sentence 2:** “A woman has been playing the violin.”

**Task:** Write two sentences that are on completely different topics.

**Sentence 1:** “A man is playing a flute.”

**Sentence 2:** “A woman is walking down the street.”

	Model	UD	STS12	STS13	STS14	STS15	STS16	STSb	SICK	Avg.
sup.	InferSent, Glove	–	52.86	66.75	62.15	72.77	66.87	68.03	65.65	65.01
	USE	–	64.49	67.80	64.61	76.83	73.18	74.92	76.69	71.22
	S-BERT (base)	–	70.97	76.53	73.19	79.09	74.30	77.03	72.91	74.89
	S-RoBERTa (base)	–	<u>71.54</u>	72.49	<u>70.80</u>	78.74	73.69	77.77	<u>74.46</u>	74.21
unsup.	Avg. GloVe	–	55.14	70.66	59.73	68.25	63.66	58.02	53.76	61.32
	Avg. BERT	–	38.78	57.98	57.98	63.15	61.06	46.35	58.40	54.81
	BERT CLS	–	20.16	30.01	20.09	36.88	38.08	16.50	42.63	29.19
	Zhang et al. (2020)	NLI	56.77	69.24	61.21	75.23	70.16	69.21	64.25	66.58
	Li et al. (2020)	NLI	59.54	64.69	64.66	72.92	71.84	58.56	65.44	65.38
	Li et al. (2020)	STS	63.48	72.14	68.42	73.77	75.37	70.72	63.11	69.57
	DINO (STS-  -x <sub>1</sub> x <sub>2</sub> )	–	64.87	78.30	66.38	79.60	76.47	76.51	<b>74.26</b>	73.77
	DINO (STS-  -x <sub>2</sub> )	STS	<b>70.27</b>	<u>81.26</u>	<b>71.25</b>	<u>80.49</u>	<u>77.18</u>	<u>77.82</u>	68.09	<b>75.20</b>

# APPENDIX

Example	English Translation	Label
P: 저는, 그냥 알아내려고 거기 있었어요. H: 이해하려고 노력하고 있었어요.	I was just there just trying to figure it out. I was trying to understand.	Entailment
P: 저는, 그냥 알아내려고 거기 있었어요. H: 나는 처음부터 그것을 잘 이해했다.	I was just there just trying to figure it out. I understood it well from the beginning.	Contradiction
P: 저는, 그냥 알아내려고 거기 있었어요. H: 나는 돈이 어디로 갔는지 이해하려고 했어요.	I was just there just trying to figure it out. I was trying to understand where the money went.	Neutral

context string	question string
1839년 바그너는 고테의 파우스트를 처음 읽고 그 내용에 마음이 끌려 이를 소재로 해서 하나의 교향곡을 쓰려는 뜻을 갖는다. 이 시기 바그너는 1838년에 빛 독촉으로 산전수전을 다 겪은 상황이라 좌절과 실망에 가득했으며 메피스토펠레스를 만나는 파우스트의 심경에 공감했다고 한다. 또한 파리에서 아브네크의 지휘로 파리 음악원 관현악단이 연주하는 베토벤의 교향곡 9번을 듣고 깊은 감명을 받았는데, 이것이 이듬해 1월에 파우스트의 서곡으로 쓰여진 이 작품에 조금이라도 영향을 끼쳤으리라는 것은 의심할 여지가 없다. 여기의 라단조 조성의 경우에 두 그의 저기에 전혀 어느 것	바그너는 고�ete의 파우스트를 읽고 무엇을 쓰고자 했는가?

**KorNLI**

**KorQuAD**